Department HW

Series 25

Piece 37 (COPY)

N 523

HW 25/37

CCR
NO. 150(1)

CONFIDENTIAL

THE APPLICATIONS OF PROBABILITY TO CRYPTOGRAPHY

by A.M. Turing

THE APPLICATIONS OF PROBABILITY TO CRYPTOGRAPHY

The theory of probability may be used in cryptography with most effect when the type of cipher used is already fully understood, and it only remains to find the actual keys. It is of rather less value when one is trying to diagnose the type of cipher, but if definite rival theories about the type of cipher are suggested it may be used to decide between them.

Meaning of probability and odds.

I shall not attempt to give a systematic account of the theory of probability, but it may be worth while to define shortly 'probability' and 'odds'. The probability of an event on certain evidence is the proportion of cases in which that event may be expected to happen given that evidence. For instance if it is known that 20% of men live to the age of 70, then knowing of Hitler only 'Hitler is a man' we can say that the probability of Hitler living to the age of 70 is 0.2 . Suppose however that we know that 'Hitler is now of age 52' the probability will be quite different, say 0.5, because 50% of men of 52 live to 70.

The 'odds' of an event happening is the ratio $p/1-p$ where p is the probability of it happening. This terminology is connected with the common phraseology 'odds of 5:2 on' meaning in our terminology that the odds are 5/2.

Probabilities based on part of the evidence

When the whole evidence about some event is taken into account it may be extremely difficult to estimate the probability of the event, even ~~xxxxx~~ very approximately, and it may be better to form an estimate based on a part of the evidence, so that the probability may be more easily calculated. This happens in cryptography in a very obvious way. The whole evidence when we are trying to solve a cipher is the complete traffic , and the events in question are the different possible keys, and functions of the keys. Unless the traffic is very small indeed the theoretical answer to the problem 'what are the probabilities of the various keys ?' will be of the form ' The key ... has a probability differing almost imperceptibly from 1 (certainty) and the other keys are virtually impossible'. But ~~xxxxxxxxx~~ ~~xxxxxxxxxxxxxxxxx~~ a direct attempt to determine these probabilities would obviously not be a practical method.

A priori probabilities

The evidence concerning the possibility of an event occurring usually divides into a part about which statistics are available, or some mathematical method can be applied, and a less definite part about which one can only use one's judgment. Suppose for example that a new kind of traffic has turned up and that only three messages are available. Each message has the letter V in the 17th place and G in the 18th place. We want to know the probability that it is a general rule that we should find V and G in these places. We first have to decide how probable it is that a cipher would have such a rule, and as regards this one can probably only guess, and my guess would be about 1/5,000,000 .

This judgment is not entirely a guess; some rather inaccurate mathematical reasoning has gone into it, something like this:-

The chance of there being a rule that two consecutive letters somewhere after the 10th should have certain fixed values seems to be about 1/500 (this is a complete guess). The chance of the letters being the 17th and 18th is about 1/15 (another guess, but not quite so much in the air). The probability of the letters being V and G is 1/676 (hardly a guess at all, but expressing a judgment that there is no special virtue in the bigramme VG). Hence the chance is 1/ 500x15x676 or about 1/5,000,000 . This is however all so vague, that it is more usual to make the judgment '1/5,000,000' without explanation.

The question as to what is the chance of having a rule of this kind might of course be solved by statistics of some kind, but there is no point in having this very accurate, and of course the experience of the cryptographer itself forms a kind of statistics.

The remainder of the problem is then solved quite mathematically. Let us consider a large number of ciphers 'chosen at random', N of them say. Of these N/5,000,000 of them will have the rule in question, and the remainder not. Now if we had three messages of each of the ciphers before us, we should find that for each of the ciphers with the rule, the three messages have VG in the required place, but of the remaining 4,999,999 N/5,000,000 only a proportion $1/676^3$ will have them. Rejecting the ciphers which have not the required characteristic we are left with

N/5,000,000 cases where the rule holds, and 4,999,999 N/5,000,000 x 676^3 cases where it does not. This selection of ciphers is a random selection of ones which have all the known characteristics of the one in question, and therefore the odds in favour of the rule holding are N/5,000,000 : 4,999,999N/5,000,000 x 676^3 i.e. 676^3 : 4,999,999 or about 60;1 on.

It should be noticed that the whole argument is to some extent fallacious, as it is assumed that there are only two possibilities, viz. that either VG must always occur in that position, or else that the letters in the 17th and 18th positions are wholly random. There are however many other possibilities worth consideration, e.g.

On the day in question we have VG in the position in question. On another day we have some other fixed pair of letters. Or

In the position 17,18 we have to have one of the four combinations VG, RH, OM, IL and by chance Vg has been chosen for all the three messages we have had. Or

The cipher is a simple substitution and VG is the substitute of some common bigramme, say TH.

The possibilities are of course endless, and it is therefore always necessary to bear in mind the possibility of there being other theories not yet suggested.

The a priori probability sometimes has to be estimated as above by some sort of guesswork, but often the situation is more satisfactory. Suppose for example that we know that a certain cipher is a simple substitution, the keys having no specially noticeable properties. Suppose also that we have 50 letters of such a message including five occurrences of P. We want to know how probable it is that P is the substitute of E. As before we have to answer two questions. How likely is it that P would

be the substitute of E neglecting the evidence of the five Es occurring in the message. Secondly 'How likely are we to get 5 Ps (a) if P is not the substitute of E (b) if P is the substitute of E. I will not attempt to answer the second question for the present. The answer to the first is simply that the probability of any letter being the substitute of E is independent of what the letter is, and is therefore always 1/26, in particular it is 1/26 for the letter P. The only guess-work here is the judgment that the keys are chosen at random.

The Factor Principle.

Nearly all applications of probability to cryptography depend on the 'factor principle' (or Bayes' theorem). This principle may first be illustrated by a simple example. Suppose that one man in five dies of heart failure, and that of men who die of heart failure two in three die in their beds, but of men who die from other causes only one in four die in their beds. (My facts are no doubt hopelessly inaccurate). Now suppose we know that a certain man died in his bed. What is the probability that he died of heart failure? Of all men numbering N say, we find that

Nx (1/5)x(2/3) die in their beds of heart failure
Nx (1/5)x(1/3) ... elsewhere ....
Nx (4/5)x(1/4) die in their beds from other causes
Nx (4/5)x(3/4) ... elsewhere ....

Now as our man died in his bed we do not need to consider the cases of men who did not die in their beds, and these consist of Nx (1/5)x(2/3) cases of heart failure and Nx (4/5)x (1/4) from other causes, and therefore the odds are 1x (2/3): 4x (1/4) in favour of heart failure. If this had been done algebraically the result would have been

A posteriori ~~probability~~ odds of the theory

= A priori ~~probability~~ odds of the theory x

$$\times \frac{\text{Probability of the data being fulfilled if the theory is true}}{\text{Probability of the data being fulfilled if the theory is false}}$$

In this ~~theory~~ 'theory' is that the man died of heart failure, and the 'data' is that he died in his bed. The general formula above will be described as the 'factor principle', the ratio

$$\frac{\text{Probability of the data if theory true}}{\text{Probability of the data if theory false}}$$ is called the factor for the theory ~~if it is~~ on account of the data.

Decibanage.

Usually when ~~we~~ we are estimating the probability of a theory there will be several independent pieces of evidence e.g. following our last example, where we want to know whether a certain man died of heart failure or not, we may know

a) He died in his bed

b) His father died of heart failure

c) His bedroom was on the ground floor

and also have statistics telling us

2/3 of men who die of heart failure die in their beds

2/5 . . . . have fathers who died of heart failure

1/2 . . . . have bedrooms on the ground floor

1/4 of men who died from other causes die in their beds

1/6 . . . . have fathers who died of heart failure

1/20 of men who die of other causes have their bedrooms on the ground floor

~~If we~~ Let us suppose that the three pieces of evidence are independent of one another ~~(i.e.) that~~ if we know that he died of heart failure, and also if we know he did not die of heart failure. That is to say we suppose ~~that~~ for instance that knowing ~~he died of heart failure, and that~~ that he slept on the ground floor does not make it any more likely that ~~t~~he died in his bed if we knew all along that he died of heart failure. When we make these assumptions the probability of a man who died of heart failure satisfying all three conditions is obtained simply by multiplication, and is (2/3)x(2/5)x(1/2), and likewise for those who died from other causes the probability is (1/4)x(1/6)x(1/20), and the factor in favour of the heart failure theory is

$$\frac{(2/3)\times(2/5)\times(1/2)}{(1/4)\times(1/6)\times(1/20)}$$

We may regard this as the product of the three factors (2/3)/(1/4) and (2/5)/(1/6) and (1/2)/(1/20) arising from the three independent pieces of evidence. Products like this arise very frequently, and sometimes one will get products involving thousands of factors, and large groups of these factors may be equal. We naturally therefore work in terms of the logarithms of the factors. The logarithm of the factor, taken to the base $10^{1/10}$ is called the 'decibanage in favour of the theory'. ~~The~~ A 'deciban' is a unit of evidence; a piece of evidence is worth a deciban if it increases the odds of the theory in the ratio $10^{1/10}$: 1 . The deciban is used as a more convenient unit than the 'ban'. The terminology was introduced in honour of the famous town of Banbury.

Using this terminology we might say that the fact that our man died in bed scores 4.3 decibans in favour of the heart failure theory ($10\log(8/3) = 4.3$). We score a further 3.8 decibans for his father dying of heart failure, and 10 for his having his bedroom on the ground floor, totalling 18.1 decibans. We then bring in the a priori odds 1/4 or $10^{-6/10}$ and the result is that the odds are $10^{12.1/10}$, or as we may say '12.1 decibans up on evens'. This means about 16;1 on.

Chapter II. Straightforward cryptographic problems.

Vigenère.

The factor principle can be applied to the solution of a Vigenère problem with great effect. I will assume here that the period of the cipher has already been determined. Probability theory may be applied to this part of the problem also, but that is not so elementary. Suppose our cipher, written out in its correct period is

D	[illegible]	[illegible]	H	S	X	[illegible]	[illegible]	[illegible]	
K	C	V	X	U	H	T	L	[illegible]	[illegible]
X	H	P	U	L	F	P	S	B	K
T	[illegible]	[illegible]	J	A	G	D	Y	O	J
T	B	W	C	Y	D	B	H	G	A
P	Z	H	C	X	L	E	Y	A	T
B	C	K	B	U	B	F	I	X	H
[illegible]	[illegible]	[illegible]	C	L	J	[illegible]	U	L	F
T	E	L	Y	F	H	L	R	T	C

Fig 1. Vigenère problem.

(It is only by chance that it makes a rectangular array).
Let us try to find the key for the first column, and for the moment let us only take into account the evidence afforded by the first letter D. Let us first consider the key B. The factor principle tells us

$$\text{Odds in favour of key B} = \text{A priori odds in favour of key B} \times \frac{\text{Probability of getting D in cipher if key is B}}{\text{Probability of getting D in cipher if key is not B}}$$

Now the a priori odds in favour of key B may be taken as 1/25. The probability of getting D in the cipher with the key B is just the probability of getting C in the clear which (using the count on 1000 letters in Fig 2) is 0.021. If however the key is not B we can have any letter other than C in the clear, and the probability is (1- 0.021)/25. Using the evidence of the D then the odds in favour of the key B are $\frac{1}{25} \cdot \frac{25 \times 0.021}{1 - 0.021}$

or [illegible] We then consider the effect of the next letter in the column R which gives a further factor of $25 \times 0.064/(1-0.064)$. We are here assuming that the evidence of the R is independent of the evidence of the B. This is not quite correct, but is a useful approximation; a more accurate method of calculation will be given later. Let us write p_α for the frequency of the letter α in plain language. Then our final estimate for the odds in favour of key B is

$$\frac{1}{25} \prod_i \frac{25\, p_{\alpha_i - 1}}{1 - p_{\alpha_i - 1}}$$

where $\alpha_1, \alpha_2, \ldots$ is the series of letters in the 1st column, and we use letters and numbers interchangeably, A meaning 1, B meaning 2, ..., Z meaning 26 or 0. More generally for key β the odds are

$$\frac{1}{25} \prod_i \frac{25\, p_{\alpha_i - \beta + 1}}{1 - p_{\alpha_i - \beta + 1}}$$

The value of this can be calculated by having a table of the decibanages corresponding to the factors $25 p_\alpha / 1 - p_\alpha$. One then decodes the column with the various possible keys, looks up the decibanages, and adds them up.

The most convenient form for doing this is a table of values of $20 \log_{10} 25 p_\alpha / 1 - p_\alpha$, taken to the nearest integer (or as we may say, the values of the score in 'half decibans'). One may also have columns showing multiples of these, and the table made of double height (Fig 3). For the first column with key B the decoded column is CQWS..OAV and we score -5 for C, -26 for Q,

A	84
B	23
C	21
D	46
E	116
F	20
G	25
H	49
I	76
J	2
K	5
L	38
M	34
N	66
O	66
P	15
Q	2
R	64
S	73
T	81
U	19
V	11
W	21
X	16
Y	24
Z	3

The value for X has been taken more or less at random ~~to allow for~~ as a compromise between real language & telegraphese. Also I added to each entry (see p).

Fig 2. Count on 1000 letters English text.

31	26	20	13	7	A
-23	-18	-14	-9	-5	B
-26	-21	-16	-10	-5	C
7	6	4	3	1	D
48	38	29	19	10	E
-28	-22	-17	-11	-6	F
-19	-15	-11	-8	-4	G
10	8	6	4	2	H
29	23	17	12	6	I
-131	-103	-77	-52	-26	J
-99	-79	-59	-40	-20	K
-2	-2	-1	-1	0	L
-6	-5	-4	-2	-1	M
23	18	14	9	5	N
23	18	14	9	5	O
-41	-33	-25	-16	-8	P
-131	-103	-77	-52	-26	Q
22	18	13	9	4	R
28	22	17	11	6	S
32	26	19	13	6	T
-31	-25	-19	-12	-6	U
-54	-43	-32	-22	-10	V
-26	-21	-16	-10	-5	W
-38	-30	-23	-15	-8	X
-20	-16	-12	-8	-4	Y
-111	-89	-67	-44	-22	Z
+31	26	20	13	7 →	A
-23	-18	-14	-9	-5	B
-26	-21	-16	-10	-5	C
7	6	4	3	1	D
48	38	29	19	10	E
-28	-22	-17	-11	-6	F
-19	-15	-11	-8	-4	G
10	8	6	4	2	H
29	23	17	12	6	I
-131	-103	-77	-52	-26	J
-99	-79	-59	-40	-20	K
-2	-2	-1	-1	0	L
-6	-5	-4	-2	-1	M
23	18	14	9	5	N
23	18	14	9	5	O
-41	-33	-25	-16	-8	P
-131	-103	-77	-52	-26	Q
22	18	13	9	4	R
28	22	17	11	6	S
32	26	19	13	6	T
-31	-25	-19	-12	-6	U
-54	-43	-32	-22	-10	V
-26	-21	-16	-10	-5	W
[illegible]	[illegible]	[illegible]	[illegible]	-8	X

Fig 3. Table for scoring a Vigenère. In units of half a deciban

September 6th B

3	2·0 ·7 · 1 0·3 0·3	3 7 0·2 ·4 2 0 0·3	2 0 0 3 ·3 [illegible]	·3 0 · 4 0·3	1·4 0·3 ·0 1·4 ·7	O [illegible] 7
3	3·7 ·7 4 3·1	5·1 1·4 3·7 ·3 5·4	·3 2 4· [illegible]	[illegible]	3·7 4· [illegible]	J. [illegible]
3	1·1 1·0 1·6 5·2 1 1	2·5 [illegible] 5·3	3·3 5·0 4·3 5·0 4·3	·7 1·1 5·0 5·0	4·3 5·0 ·8 1·4 6·7	K [illegible] 4
B	1 4 0 2 7 5·7 2·6	·8 2·6 ·2 5 2 2·2	1·2 1·2 4·1 ·1 ·2	1·0 3 6 ·3 6 4 2 3	·1 ·1 7 3 3 ·2	F [illegible] 4
B	6 0 ·4·9 2·6 6·0	3·6 4 7 2 2 2·6 1·4	1 8 3·8 2 2·1 4 1	5·4 ·7 ·1 3 8 7·1	3 7 6 0 6 7 3 3·2	R 264 - 179 11
B	3·9 3·7 4 0 2·7 4·1	2·4 3·7 3 7 5 6 9	4·1 2·5 1 2 3 1 3	·8 ·4 ·6 ·4	1·4·6 1·4 1·8 1 1	V [illegible] - 147 6
3	·· 1 2 3·0 5 2 3·5	1 · 2 2 = 1 5 3	·6 1 · 3·9 2·2 ·5	5 2 2 2 1 8 3·5	3 5 1 9 2·9 7 9 1 5	M 120 - 95 4
3	3·7 2·0 2·0 2·0	3 5 4 3 7 3 7 2 0	3·7 5 4 3 7 1 ·3	1 1 5 4 ·3 2·0	5 4 3 7 [illegible]	T [illegible] - 63 7

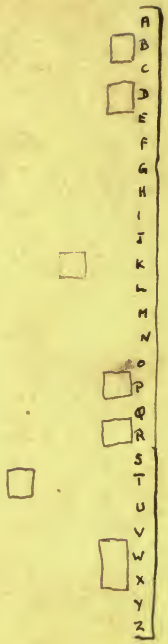

Fig 4. Apparatus for scoring in Vigenère. Pencil marks arranged for 1st column of Fig 1.

D	K	Q	H	S	H	Z	M	N	P
R	C	V	X	U	H	T	E	A	Q
X	H	P	U	E	P	P	S	B	K
T	W	U	J	A	G	D	Y	O	J
T	H	W	C	Y	D	Z	H	G	A
P	Z	K	O	X	O	E	Y	A	E
B	O	K	B	U	B	P	I	K	R
W	W	A	C	E	J	P	H	L	P
T	U	Z	Y	F	H	L	R	Y	C

Scores for various keys

	P	O	IC	QU
A	-27	X	X	X
B	-17	X	X	-2
C	X	3	16	X
D	X	9	9	X
E	-6	X	X	X
F	X	X	X	X
G	X	X	X	-3
H	X	1	-3	X
I	X	X	17	X
J	X	X	X	13
K	X	X	X	-15
L	2	X	X	X
M	X	X	X	X
N	X	X	X	X
O	X	28	X	X
P	43	X	X	X
Q	X	X	X	22
R	X	X	X	X
S	X	X	-6	X
T	8	X	-15	X
U	X	X	X	22
V	X	X	X	X
W	X	16	[illegible]	X
X	X	X	-15	X
Y	X	X	-15	X
Z	X	-13	X	X

Best keys: P O IC QU

Possible decodes

O	W	IO	RN	G
C	O	NT	HD	I
I	T	HN	EA	S
E	I	MW	TT	O
E	T	OU	MI	M
A	L	CI	YU	L
M	A	CI	LH	I
H	[illegible]	SY	MI	S

Fig K Scoring and solving a Vigenère.

-5 for ~~E~~ W, 17 for the three letters S, 5 for O, 2 for A and -10 for V, totalling -17. These calculations can be done very quickly by the use of the transparent gadget Fig 4, in which squares are ringed in pencil to show the number of letters occurring in the column. The gadget may be placed over Fig 3 in various positions corresponding to the various possible keys. The score is obtained by adding up the numbers showing through the various squares. In Fig 5 the ~~possible~~ alphabet has been written in a vertical column below the cipher text of Fig 1, each letter representing a possible key. The score for each key has been written opposite the key, and under the relevant column. An X denotes a bad score, not worth adding up. Usually there will be -15 or worse. It will be seen that for the first column P, having a score of 43 is extremely likely to be right, especially as there is no other score better than 8. If we neglect this latter fact the odds for the key are (1/25) $10^{2.15}$ i.e. about 5:1 on. The effect of decoding this column with key P has been shown underneath. For the second column the best key is G, but is by no means so certain as the first column. The decode for this column is also shown, and provides very satisfactory combinations with the first column, confirming both the keys. (This confirmation could also be based on probability theory, given a table of bigramme frequencies). In the third column I and C are best although D would be very possible, and in the fourth column Y and U are best. Writing down the possible decodes we see that the first line

must read D.ING and this makes the other lines read CONDI, ITHAS, [illegible], ETOT ,ALCOL, [illegible], [illegible],EGRET. By filling in the word 'conditions' the whole can now be decoded.

A more accurate argument would run as follows. For the first column, instead of setting up as rival theories the two possibilities that B is the key and that B is not the key we can set up 26 rival theories that the key is A or B or ... or Z, and we may apply the factor principle in the form:-

$$\frac{\text{A posteriori probability of key A}}{\text{A priori probability of key A} \times \text{Probability of getting the given column with key A}}$$

$$= \frac{\text{A posteriori probability of key B}}{\text{A priori probability of key B} \times \text{Probability of getting the given column with key B}}$$

= etc.

The argument to justify this form of factor principle is really the same as for the original form. Let q_β be the a priori probability of key β. Then out of N cases we have Nq_β cases of key β. Let $P(\beta,C)$ be the probability of getting the ~~key~~ column C with key β, then when we have rejected the cases where we get columns other than C we find that there are $Nq_\beta P(\beta,C)$ cases of key β left, i.e. the a posteriori probability of key β is $Kq_\beta P(\beta,C)$ where K is independent of β.

We have therefore to calculate the probability of getting the column C with key β and this is simply $\prod_i P_{x_i-\beta+1}$ i.e. the product of the frequencies of the decode letters which we get if the key is β.

Since the a priori probabilities of the keys are all equal we may say that the a posteriori probabilities are in the ratio $\prod_i P_{\alpha_i - \beta + 1}$, i.e. in the ratio $\prod_i (26 P_{\alpha_i - \beta + 1})$ which is more convenient for calculation. The final value for the probability is then

$$\prod_i (26 P_{\alpha_i - \beta + 1}) \Big/ \sum_\beta \prod_i (26 P_{\alpha_i - \beta + 1})$$

The calculation of the products $\prod_i (26 P_{\alpha_i - \beta + 1})$ may be done by the method recommended before for $\prod_i 25 P_{\alpha_i - \beta + 1} / 1 - P_{\alpha_i - \beta + 1}$ (The table in Fig 3 was in fact made up for $\prod_i (26 P_{\alpha_i - \beta + 1})$. The differences between the two tables would of course be rather slight). The new result is more accurate than the old because of the independence assumption in the original result.

If we only want to know the ratios of the probabilities of the various keys there is no need to calculate the denominator $\sum_\beta \prod_i (26 P_{\alpha_i - \beta + 1})$. This denominator has however another importance: it gives us some evidence about our other assumptions, such as that the cipher is Vigenere, and that the period is 10. This aspect will be dealt with later (p.   ).

A letter subtractor problem

A substitution with period 91 x 95 x 99 is obtained by superimposing three substitutions of periods 91, 95, and 99, each substitution being a Vigenere composed of slides of 0,1,2,3,4,5,6,7,8, or 9. The three substitutions are known in detail, but we do not know for any given message at what point in the complete substitution to begin. For many messages however we can provide a more or less probable crib. How can we test the probability of a crib before attempting to solve it? It may be assumed that approximately equal numbers of slides 0,1,... 9 occur in each substitution.

The principle of the calculation is that owing to the way in which the substitution is built up, not all slides are equally frequent, e.g. a slide of 25 can only be the sum of slides of 9,8 and 8 or of 9,9 and 7 whilst a slide of 15 can be any of the following

9,6,0	8,7,0	7,7,1	6,6,3
9,5,1	8,6,1	7,6,2	6,5,4
9,4,2	8,5,2	7,5,3	
9,3,3	8,4,3	7,4,4	

A crib will therefore, other things being equal, be more likely if it requires a slide of 15 than if it requires a slide of 25. The problem is to make the best use of this principle, by determining the probability of the crib with reasonable accuracy, but without spending long over it.

We have to find out the probability of getting a given slide. To do this we can apply several methods.

(a) We can produce a long stretch of ~~xx~~ key by addition and take a count of the resulting slides. This is obviously a very general method, and requires no special mathematical technique. It may be rather laborious, but by interpreting a small count with common sense one can probably get quite good results.

(b) There are 1000 possible combinations of slides all equally likely, viz 000,001,...,999 . ~~Ifxxx~~ We can add up the digits in these and ~~xxx~~ take the remainder on division by 26, and then count the number of combinations giving each of the possible remainders.

(c) We can make use of a trick which might appear to be rather special, but is really applicable to a multitude of problems. Consider the expression

$$f(x) = (1 + x + x^2 + \dots + x^9)^3$$

For each possible way of expressing a number n as the sum of three numbers 0,...,9, say $n = m_1 + m_2 + m_3$ there is a term $x^{m_1} x^{m_2} x^{m_3}$ in $f(x)$, x^{m_1} coming out of the first factor, x^{m_2} out of the second, and x^{m_3} out of the third. Hence the number of ways of expressing n in the form $n = m_1 + m_2 + m_3$ is the coefficient of x^n in $f(x)$ i.e. in

$$\frac{(1 - x^{10})^3}{(1 - x)^3}$$

or in

$$(1 - 3x^{10} + 3x^{20} - x^{30})(1 - x)^{-3}$$

Expanding $(1-x)^{-3}$ by the binomial theorem

$$(1-x)^{-3} = 1+3x+6x^2+10x^3+15x^4+21x^5+28x^6+36x^7+$$
$$+45x^8+55x^9+66x^{10}+78x^{11}+91x^{12}+105x^{13}+120x^{14}+136x^{15}$$
$$+153x^{16}+171x^{17}+190x^{18}+210x^{19}+231x^{20}+253x^{21}+276x^{22}+300x^{23}$$
$$+325x^{24}+351x^{25}+378x^{26}+406^{27}+435x^{28}+\ldots$$

Now multiply by $1-3x^{10}+3x^{20}-x^{30}$ and we get

$$f(x) = 1+3x+6x^2+10x^3+15x^4+21x^5+28x^6+36x^7+45x^8+55x^9+$$
$$+63x^{10}+69x^{11}+73x^{12}+75x^{13}+75x^{14}+73x^{15}+69x^{16}+63x^{17}+$$
$$+55x^{18}+45x^{19}+36x^{20}+28x^{21}+21x^{22}+15x^{23}+10x^{24}+6x^{25}+3x^{26}+x^{27}$$

This means to say that the chances of getting totals of 0,1,2,... are in the ratio 1, 3, 6, 10,... The chances of getting remainders of 0,1,2,... on division by 26 are in the ratio 4, 4, 6, 10, 15, ... To get true probabilities these must be divided by their total which is conveniently 1000.

(d) There are two other methods, both connected with the last method but ~~xxxxxxx~~ not relying so much on the special ~~x~~ features of the problem. They will be discussed later.

Suppose then that the probabilities have been calculated by one method or the other(as in fact ~~xxxxxxxxxxxxxfor~~ we have done under (c)). We can then estimate the values of cribs. Let us suppose that a possible crib for a message beginning MVHWUSXOWBVMMK was ~~XXXXXXXXXXXX~~ AMBASSADOR so that the slides were 12, 9, 6, 22, 2, 0, 23, 11, 14, The slide of 12 gives us some slight evidence in favour of the crib being right for

slides of 12 occur with frequency 0.073 with right cribs, whilst with wrong cribs they occur with frequency only 1/26. The factor in favour of the crib is therefore 26x0.073 or about 1.9 . A similar calculation may be made for each of the slides, but of course the work may be greatly speeded up by having the values of the factors $26\ C_s/1000$ in half decibans tabulated: here C_s is the coefficient of x^s in the above polynomial $f(x)$. The table is given below (Fig 6)

1	0	-20
2	25	-16
3	24	-12
4	23	-8
5	22	-6
6	21	-3
7	20	-1
8	19	1
9	18	3
10	17	4
11	16	5
12	15	6
13	14	6

Fig 6. Scores in half decibans of the various slides.

Evaluating this crib by means of this table we ~~xxx~~ score

$$6+3-3-6-16-20-8+5+6\ (=-33)$$

i.e. the crib is worse by a factor of $10^{-33/20}$ than it was before e.g. if the a priori odds of the crib were 2:1 against it becomes 98:1 against. This crib was in fact made up

at random i.e. the letters of the cipher text were chosen at random. Now let us take one made up correctly, i.e. really enciphered by the method in question, but with a random chosen key.

```
 N Y X L N X I   H H
 A M B A S S A D O R
13  22  21  8  19
   12 11  5  13  16   (slides)
```

This scores 15 so that if it were originally 2:1 against it now becomes nearly 3:1 on.

Having decided on a crib the natural way to test it is to have a catalogue of the positions in which a given series of slides is obtained if the 91 period component is omitted. We make 91 different hypotheses as to this third component, draw an inference as to what is the part of the slide arising from the components of periods 95 and 9? combined. This we look up in the catalogue. This process is fairly lengthy, and as the scoring of the crib takes only a minute it is certainly worth doing.

Theory of repeats

Suppose we have a cipher in which there are several very long series of substitutions which can be used for enciphering a message, but that one may sometimes get two messages enciphered with the same series of substitutions (or more truly, the series of substitutions for one message being those for another with some at the beginning omitted). In such a case let us say that the messages 'fit', or that they fit at such and such a distance, the distance being the number of substitutions which have to be omitted from the one series to obtain the other series. One will frequently want to know whether two messages fit or not, and we may find some evidence about this by examining the repeats between them. By the repeats between them I mean this. One writes out the cipher texts of the two messages with the letters which are thought to have been enciphered with the same substitution under one another. One then writes under these messages a series of letters o and x, an o being written where the cipher texts differ and an x where they agree. These series of letters o and x will begin where the second message begins and end where the first to end ends. ~~To complete the information about the repetition figure~~ This series of letters o and x may be called the repetition figure. It may be completed by adding at the end an indication of how many letters there are which do not overlap, and which message they belong to. As an example

GFRLIKQGVBMILAFIXKMOROGBYSKYXDAZCHQUMRKBZLDLDOHCAVTIPRSD

VLOVDYQCLJSOPYGBMBKYXDAZMBFIOPTFCXDOD

^{8}xooooooooooxooxxooxxxxxxooooooooooxox^{11}

On the whole one expects that a fit is more likely to be right the more letters x there are in the repetition figure, and that long series of letters x are especially desirable. This is because ~~xxy~~ it would not be very unusual for two fairly common words to lie directly under one another when the clear texts are written out, thus

THEMAINCONVOYWILLARRIVE . . .
    ALLCONVOYSMUSTREPORT. . .
    xooxxxxxxoooooxoooo . . .

If the corresponding cipher texts really fit, i.e. if the letters in the same column are enciphered with the same substitution, then the condition for an x in the repetition figure of the cipher texts is that there be an x in the repetition figure of the corresponding clear text. Now series of several consecutive letters x can occur quite easily as above by two ~~xo~~ identical words coming under one another, or by such combinations as

ITISEASIERTOTEACHTHANALGEBRA . . .
  THERAINWASSUCHTHATHECOULD . . .
  ooooooooooooxxxxxoooooooo . . .

if the messages really fit, but if not they can only occur by complete coincidence. One therefore tends to believe that there is a fit when one gets such series of letters~~x~~ x . As regards single cases of x the value of them is not so clear, but one can see that if p_α is the frequency of letters α in ~~clear~~ plain language then the frequency of letters x as a whole in comparisons of plain language with plain language is $\sum_\alpha p_\alpha^2$, whilst for wrong fits of cipher text it is 1/26 which is necessarily less. Given

a sufficiently long repetition figure one should therefore be able to tell whether it is a fit or not simply by counting the letters x and o.

So much is well known. The real point of this section is to show how these ideas can be developed into an accurate method of estimating the probabilities of fits.

Simple form of theory. The complete theory takes account of the various possible lengths of repeat. As this theory is somewhat complicated it will be as well to give first two simplified forms of the theory. In both cases the simplification arises by neglecting a part of the evidence. In the first simplified form of theory we neglect all evidence except the number of letters x and the number of letters o. In the other simplified form the evidence is the number of series of (say) four consecutive letters x in a repetition figure.

When our evidence is just the number of times x occurs (n let us say) in the repetition figure, and the length of the repetition figure (N say), then the factor in favour of the fit is

$$\frac{\text{Probability of a right repetition figure of length N having n occurences of x}}{\text{Probability of a wrong repetition figure of length N having n occurrences of x}}$$

As an approximation we may assume that the numerator of this expression has the same value as if the right repetition figures were produced letter by letter by independent random choices, with a certain fixed probability of getting an x at each stage. This probability will have to be $\beta = \sum_\alpha p_\alpha^2$. The numerator

is then

(Number of repetition patterns with length N and n occurrences of x)

times (Probability of getting a given such repetition pattern by the process just mentioned)

which we may write as R(N,n) Q(N,n). Now let us denote by y_i the i th symbol of the given repetition pattern, and put $\tau_x = \beta$ and $\tau_o = 1-\beta$. Then $\varphi(N,n)$, the probability of getting the repetition pattern is $\prod_{i=1}^{N} \tau_{y_i}$ which simplifies to $\beta^n (1-\beta)^{N-n}$. We may do a similar calculation for the denominator, but here we must take $\beta = \frac{1}{26}$ since all letters occur equally frequently in the cipher. The denominator is then $R(N,n)\left(\frac{1}{26}\right)^n\left(\frac{25}{26}\right)^{N-n}$. In dividing to find the factor for the fit $R(N,n)$ cancels out, leaving $(26\beta)^n\left(\frac{26}{25}(1-\beta)\right)^{N-n}$. In other words we score a factor of 26β for an x and a factor of $\frac{26}{25}(1-\beta)$ for an o. More convenient is to regard it as $10\log_{10} 25\beta/1-\beta$ decibans for an x and $10\log_{10}\frac{26}{25}(1-\beta)$ decibans per unit length of repetition figure ('per unit overlap').

An alternative argument, leading to the same result, runs as follows. Having decided to neglect all evidence except the overlap and the number of repeats we pretend that nothing else matters, i.e. that the form of the figure is irrelevant. In this case we can regard each letter of the repetition figure as independent evidence about the fit. If we get an x the factor for the fit is

$$\frac{\text{Probability of getting an x if the fit is right}}{\text{Probability of getting an x if the fit is wrong}}$$

i.e. $\frac{\beta}{1/26}$

Similarly the factor for an o is $\frac{1-\beta}{25/26}$

In either form of argument it is unnecessary to calculate the number $F(N,n)$. In this particular case there is no particular difficulty about it: it is the binomial coefficient. In some similar problems this cancelling out is a great boon, as we might not be able to find any simple form for the factor which cancels. The cancelling out is a normal feature of this kind of problem, and it seems quite natural that it should happen when we think of the second form of argument in which we think of the evidence as consisting of a number of independent parts.

The device of assuming, as we have done here, that the evidence which is not available is irrelevant can often be used and usually leads to good results. It is of course not supposed that the evidence really is irrelevant, but only that the ~~xolxxixx fxxtxr~~ error resulting from this assumption when used in this kind of way is likely to be small.

In the second ~~fxrx~~ simplified form of theory we take as our evidence that a particular part of the repetition figure is Oxxxxo (say, or alternatively oxxxyxo say). The factor is then

$$\frac{\text{Frequency of oxxxxo in right repetition figures}}{\text{Frequency of oxxxxo in wrong repetition figures}}$$

The denominator is $\left(\frac{1}{26}\right)^4\left(\frac{25}{26}\right)^2$ and the numerator can be estimated by taking a sample of language hexagrams and counting the number of pairs that have the repetition figure oxxxxo. The expectation of the number of such pairs is the sum for all pairs of the ~~sxpex~~ probabilities of those pairs ~~hxinx~~ having the desired repetition figure i.e. is the number of such pairs (viz $N(N-1)/2$ where N is the size of the sample) multiplied

by the frequency of oxxxo repetition figures. This frequency may therefore be obtained by division if we compare the expected number of these repetition figures to the actual number.

<u>General form of theory</u>. It is not of course possible to have statistics of every conceivable repetition figure. We must make some assumption to reduce the variety that need be considered. The following assumption is theoretically very convenient, and also appears to be a very good approximation.

<u>The probabilities of repeats at two points known to be separated by a point where there is known to be no repeat are independent</u>.

We may also assume that the probability of a repeat is independent of anything but the repetition figure in its neighbourhood. (We may however as a refinement produce different statistics for different types of messages, and different positions in a message) . We can therefore think of a repetition figure as being produced by selecting the symbols of the figure consecutively, the probability of getting an x at each stage being determined by the repetition figure from the point in question back as far as the last o. Sometimes this will take us back as far as the beginning of the message, and will include the number telling us how many more letters there are which do not repeat at all. We need in practice only distinguish two cases, where this number is 0 and when it is more. We therefore have to distinguish the following cases

We may also neglect the question as to which message comes first.

o	a_0	some	b_0	none	c_0
ox	a_1	some x	b_1	none x	c_1
oxx	a_2	some xx	b_2	none xx	c_2
oxxx	a_3	some xxx	b_3	none xxx	c_3
. . .		. . .		. . .	

The entries a_0, a_1, b_0 etc. opposite the repetition figures are the notations we are adopting for the probability of getting another x following such a figure. Strictly speaking we should also bring in a notation for the probability of the message coming to an end after any given repetition figure. As the repeats at the end of a comparison do not appear to behave very differently from those in the main part of the message I shall neglect this complication by assuming that the probability of getting an o added to the probability of getting an x is 1, and that afterwards one cuts off the end of the series arbitrarily.

Let us calculate the factor for the repeat figure

none	x	x	x	x	o	o	o	x	o	x	x	x	o	o	x	x	some
	c_0	c_1	c_2	c_3	$1-c_4$	$1-a_0$	$1-a_0$	a_0	$1-a_1$	a_0	a_1	a_2	$1-a_3$	$1-a_0$	a_0	a_1	
	$\frac{1}{26}$	$\frac{1}{26}$	$\frac{1}{26}$	$\frac{1}{26}$	$\frac{25}{26}$	$\frac{25}{26}$	$\frac{25}{26}$	$\frac{1}{26}$	$\frac{25}{26}$	$\frac{1}{26}$	$\frac{1}{26}$	$\frac{1}{26}$	$\frac{25}{26}$	$\frac{25}{20}$	$\frac{1}{26}$	$\frac{1}{26}$	

Underneath each symbol has been written the probability that one would get that symbol, knowing the ones which preceded, both for the case of a right and of a wrong repetition figure. The factor for the fit is the product of the first row divided by the product of the second. It is convenient to ~~divide~~ split this up, as indicated by the vertical lines into the product of

$$\frac{c_0 c_1 c_2 c_3 (1-c_4)}{\left(\frac{1}{26}\right)^4 \frac{25}{26}}$$

$$\frac{1-a_0}{\frac{25}{26}}$$ occurring three times

$$\frac{a_0(1-a_1)}{\frac{1}{26} \cdot \frac{25}{26}}$$

$$\frac{a_0 \, a_1 \, a_2 (1-a_3)}{\left(\frac{1}{26}\right)^3 \frac{25}{26}}$$

$$\frac{a_0 \, a_1}{\left(\frac{1}{26}\right)^2}$$

and this product may be put into the form of the product of

$$\frac{c_0 c_1 c_2 c_3 (1-c_4)}{(1/26)^4 \; 25/26} \cdot \left(\frac{1-a_0}{25/26}\right)^{-5}$$ which we call 'the factor for an initial tetragram repeat, level'

$$\frac{a_0 (1-a_1)}{\frac{1}{26} \cdot 25/26} \cdot \left(\frac{1-a_0}{25/26}\right)^{-2}$$ the factor for a single repeat

$$\frac{a_0 a_1 a_2 (1-a_3)}{\left(\frac{1}{26}\right)^3 \cdot 25/26} \cdot \left(\frac{1-a_0}{25/26}\right)^{-4}$$ the factor for a trigramme

$$\frac{1-a_0}{1-a_2}$$ the correction for a final bigramme

$$\left(\frac{1-a_0}{25/26}\right)^{16}$$ the factor for an overlap of 16.

$$\frac{a_0 \, a_1 (1-a_2)}{\left(\frac{1}{26}\right)^2 \frac{25}{26}} \cdot \left(\frac{1-a_0}{25/26}\right)^{-3}$$ the factor for a bigramme.

We shall neglect the correction for a final bi gramme (or whatever it may be). It is in any case rather small, and ~~ix~~ vanishes if the repetition figure ends with o: also with our conventions the whole question of the ends of repetition figures has been left rather in doubt.

Now let us put

$$a_0 a_1 \dots a_r (1 - a_{r+1}) = k_r \quad\text{—— } k_{r+1}$$
$$b_0 b_1 \dots b_r (1 - b_{r+1}) = j_r$$
$$c_0 c_1 \dots c_r (1 - c_{r+1}) = k_r$$

The values of the i_r can be obtained as follows. We take a number of plain language messages and leave out two or three words at the beginning. Then combine the messages to form one long message: this message may be made to 'eat its own tail' i.e. it may be written round a circle. If the message were compared with itself in every possible position, except level, we should expect to get repetition figures ~~including~~ which when divided up as above by vertical lines after each o, contain $h \frac{N(N-1)}{2} k_r$ ($= N_r$) parts which consist of r symbols x followed by an o, or as we may say N_r 'actual r-gramme repeats'. ~~This~~ The values of N_r can be calculated given ~~from~~ the 'apparent number of r-gramme repeats' M_r for each r. This apparent number of r-gramme repeats is the number of series of r consecutive symbols x in the repetition figures regardless of what precedes or follows the series. By considering the way in which an actual repeat can give rise to apparent repeats of various lengths we see that

where h is the probability of an o

$$M_r = N_r + 2N_{r+1} + 3N_{r+2} + \dots$$

and therefore

$$M_r - M_{r+1} = N_r + N_{r+1} + N_{r+2} + \dots$$

and

$$(M_r - M_{r+1}) - (M_{r+1} - M_{r+2}) = N_r$$

The calculation of j_r may perhaps best be done by comparing the beginners of a number of messages with the long circular message, and the values of i_r by comparing the beginners among themselves. A similar technique of actual and apparent numbers of repeats can be used. I shall not go into this in detail.

The formulae required may now be assembled.

μ_r : decibanage for an r-gramme repeat

ν : negative decibanage for unit overlap

$s_{\beta,r}$: number of occurrences in the statistics of the r- gramme β.

N = total number of letters in the statistics

Then if
$$M_r = \sum_{\beta} s_{\beta,r}(s_{\beta,r}-1)/2$$
$$N_r = M_r - 2M_{r+1} + M_{r+2}$$
$$L = N(N-1)/2$$
$$k_r = N_r / Lh$$

h may be calculated as follows. From the identity
$$(1-a_0) + a_0(1-a_1) + a_0 a_1 (1-a_2) + \dots = 1$$
we get
$$k_0 + k_1 + k_2 + \dots = 1$$
i.e.
$$\frac{L - M_1}{Lh} = 1$$
$$1 - a_0 = k_0 = N_0 / L - M_1 = \frac{L - 2M_1 + M_2}{L - M_1}$$
$$\mu_r = 10 \log_{10} \left(\frac{26^{r+1} k_r}{25} \right) + (r+1)\nu$$
$$\nu = -10 \log_{10} \frac{26(1-a_0)}{25}$$

Transposition ciphers

In making calculations about substitution ciphers we have often found it useful to treat the plain language as if it were produced by independent choices for the letters, using certain fixed frequencies with which the letters are chosen. Our method for Vigenere and one of the simplified forms of repeat theory could be based on this sort of assumption. With a transposition cipher however such an assumption would be useless or worse than useless, for it would result in the conclusion that all transpositions were equally likely. We have therefore to make a slightly less crude assumption, and the one which suggests itself is that the letters forming the plain language are chosen consecutively, the probability of getting a particular letter depending only on what the letter is and what the preceding letter was. It is easily verified that if $P_{\alpha\beta}$ is the proportion of bigrammes $\alpha\beta$ in plain language and P_α the frequency of the letter α then the probability $q_{\alpha\beta}$ of a letter β following an α is $P_{\alpha\beta}/P_\alpha$. The probability of a piece of plain language of length L letters saying $\alpha_1 \; \alpha_2 \dots \; \alpha_L$ is then $P_{\alpha_1} q_{\alpha_1\alpha_2} q_{\alpha_2\alpha_3} q_{\alpha_3\alpha_4} \cdots q_{\alpha_{L-1}\alpha_L}$ which may also be written as $J(\alpha_1, \dots, \alpha_L)$. We may also calculate the probability of a piece of plain language having certain given letters in given places, the remainder of the message being unspecified. The probability is given by

by

$$\Sigma(\xi_1 \ldots \xi_L \text{ consistent with data})\ J(\xi_1, \ldots, \xi_L)$$

and if the data is that the known letters are

$$\underset{n_1 \text{ dots}}{\ldots} / \beta_1 \underset{n_2 \text{ dots}}{\ldots} / \beta_2 \ldots \quad \ldots / \beta_{r-1} \underset{n_r \text{ dots}}{\ldots} / \beta_r \ldots \qquad (D)$$

it is approximately

$$p_{\beta_r} \quad \left(\prod_r \beta_r\right) \cdot \prod_{n_{r+1}=0} \frac{p_{\beta_r \beta_{r+1}}}{p_{\beta_r} p_{\beta_{r+1}}} \qquad (A)$$

A more or less rigorous deduction of this approximation from the assumptions above is given ~~below~~ at the end of this section. For the present let us see how it can be applied. If we have two theories ~~xxxxxxxxx~~ about the transposition of which the one requires the above pattern of letters, and the other brings the same letters in to positions in which no two of them are consecutive, then the factor in favour of the first as compared with the second is

$$\prod_{n_{r+1}=0} \frac{p_{\beta_r \beta_{r+1}}}{p_{\beta_r} p_{\beta_{r+1}}}$$

We can apply this straightforwardly to the case of ~~xxxxxx~~ simple transposition by columns. The following text is known to be a simple transposition of a certain type of German text with a key length of not more than 15.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A	-7	5	10	-4	-5	-7	3	-4	-25	-7	-8	-1	1	3	-20	-6	-2	-1	-6	-6	8	0	-11	-20	-7	-13	A
B	-7	6	-10	-9	9	-13	-6	-13	1	-11	-12	-7	-2	-18	4	-13	-2	1	-13	-15	-1	-16	-12	-18	-4	-12	B
C	-14	-3	-3	-18	-19	-20	-19	27	-21	-10	3	-12	-12	-21	-15	-13	-2	-14	-14	-21	-21	-5	-17	-17	-14	-17	C
D	5	-8	-18	-18	4	-6	-13	-16	-4	2	-9	-11	-6	-13	4	-4	-2	9	-6	-11	2	-2	-9	-10	-5	-4	D
E	-15	4	0	-8	-15	-5	-2	-5	10	-8	-14	1	-1	5	-22	-6	-3	8	-2	-6	-5	-6	-11	-13	-8	-4	E
F	-2	-11	-20	-9	-2	10	-3	-15	-12	4	-2	1	-8	-11	6	0	-3	-8	-8	-1	9	-8	-8	1	-2	-1	F
G	-1	-8	-18	-5	8	-10	-2	-13	-10	2	6	-3	-13	-11	-10	-14	-3	0	-3	-9	-2	-7	-7	1	2	-2	G
H	1	-10	-12	-8	4	-5	-11	-12	-2	5	-10	2	-3	-10	-2	-2	-2	4	-1	9	-11	-7	-4	-7	-15	-8	H
I	-14	-4	10	-10	0	-1	2	-18	-17	-6	-5	0	-1	9	-17	-7	-1	-10	-1	4	-19	-7	-16	-3	-9	-4	I
J	3	3	-4	1	-3	0	-1	2	-6	14	1	-3	1	-7	-3	7	-2	-12	1	-8	-4	5	-2	0	9	-12	J
K	-2	-9	-12	3	-3	1	-7	-9	-9	4	20	-2	1	-14	-15	-13	7	-6	-5	0	0	-3	-6	-17	-5	-8	K
L	6	0	-6	2	-4	-7	-1	-15	1	-3	-14	8	-4	-2	-2	-5	8	-18	-5	-5	2	-1	-10	3	-5	-3	L
M	6	-1	-17	-6	1	-9	-6	-5	5	1	-10	-14	15	-14	0	-2	-2	-14	-6	-5	-11	-4	-7	3	-6	2	M
N	-1	-8	-18	10	-6	3	11	-9	-8	2	-6	-11	-5	-7	-2	-9	2	-20	6	-6	4	-3	-6	3	0	-5	N
O	-10	-8	-10	-6	-3	4	-6	-13	-18	0	-1	2	6	0	1	2	-2	9	0	2	-18	3	-11	6	-6	-2	O
P	2	-7	-13	-13	4	-3	-14	-5	-8	3	-12	0	-2	-8	6	8	-2	5	-15	-10	5	-12	-12	-13	-11	-1	P
Q	-3	-2	-2	-2	-3	-3	-3	-2	-2	-1	-3	-2	-2	-3	-2	-2	3	-3	-3	-3	13	-2	-2	-2	-2	-2	Q
R	2	3	-3	5	2	0	-6	-10	-2	2	-1	-6	-2	-9	-6	1	-1	-11	-1	2	-1	-2	0	1	-1	2	R
S	-3	-1	11	-2	-4	-7	-13	-12	2	-6	1	-9	-10	-7	-5	8	-3	-19	5	7	-8	5	-9	-15	1	4	S
T	3	-3	-14	-7	5	-3	-11	-11	-4	1	-1	-4	-5	-9	-2	-7	-3	-2	-2	5	-5	-3	-4	2	-3	9	T
U	-11	-4	-3	-15	0	5	-6	-2	-26	-12	-16	10	-2	8	-11	3	2	-3	-3	-11	-2	-9	-11	-12	-9	-20	U
V	-13	-16	-15	-18	2	-15	-16	-5	12	-10	-7	-16	-15	-17	14	-12	-2	-17	-18	-8	-11	10	-14	-4	8	-8	V
W	-1	-17	-17	-17	4	-18	-18	-18	2	-10	-9	-13	-10	-20	21	-12	-2	-19	-20	-19	-13	-14	-3	-16	-6	-16	W
X	1	1	-13	6	1	0	-4	-13	-14	3	0	-12	2	-10	-12	-2	8	-15	-6	-4	-5	9	1	10	-6	3	X
Y	-11	-1	-14	7	-5	-4	-9	-9	-11	-9	0	-9	-6	-5	-4	-11	-2	-11	-6	-6	-7	9	-2	-13	25	-2	Y
Z	-13	-13	-17	-13	-4	-12	-8	-14	-7	-3	-10	-11	-17	-16	-1	-5	-2	-20	-13	-10	8	-8	27	0	-6	-2	Z
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	

Fig 6. Exclusive bigramme scores in half decibans, i.e. $20 \log_{10} \frac{P_{\alpha\beta}}{P_\alpha P_\beta}$ for a certain kind of German traffic.

S A T P T W S F A S T A U T E M A I E U F H W T J T D D G C
N L T S E F C U I L B O E Y Q H G T J T E E F [illegible] E O R T A R
U R N L N N N N A I E O T U S H L E S B F B R N D X G N J H
U A N W R

To solve this transposition we may try comparing the first six letters S A T P T W which we know form part of one column with each other series of six letters in the message, for we know that one such comparison will give entirely bigrammes occurring in the decode. We may try first

S F
A A
T S
P T
T A
W U
~~B W~~

The factor for a transposition which brings these letters together, as compared with one which leaves them apart is

$$\frac{p_{SF}}{p_S p_F} \cdot \frac{p_{AA}}{p_A p_A} \cdots \frac{p_{SF}}{p_S p_T} \quad \frac{p_{WU}}{p_W p_U}$$

By using a table of values of $20 \log_{10} \frac{p_{\alpha\beta}}{p_\alpha p_\beta}$ made up for the type of traffic in question, and given to the nearest integer (table of values of $\frac{p_{\alpha\beta}}{p_\alpha p_\beta}$ expressed in half-decibans) we get the product by addition. Such a table is shown in Fig 6 . The scores for this particular column are SF -7, AA -7, TS -2, PT -10, TA -3, WU -13, ~~BW -7~~, totalling ~~-29~~ -36. If we consider this combination as a priori about 100:1 against (there are 95 letters in the message) it is a posteriori about 3000:1 against. Similar scoring

may be done for every possible comparison of S A T P T W with six consecutive letters of the message. The comparison may be made both with S A T P T W as earlier and as later column; one may also use the last six letters of the message H U A N W R . The results of doing this are shown in ~~fix~~ Fig 7 . The message has been written out vertically. The first column of figures after the message gives the scores for S A T P T W as earlier column, entered against the first letter of the later column, e.g. the -36 as calculated above gets entered against the F of F A S T A U . The second column after the message ~~is~~ consists of the scores for H U A N W R as first column [and the column before the message gives the scores for H U A N W R as second column]. One of these columns has been worked out in detail but in the other two crosses have been put in where the scores are very bad. The scores which eventually turned out to be right are ringed. The fourth comparison, which did not have to be done scored very badly viz. -27. Amongst the good scores which were wrong there was one score of 37. It was not difficult to see that this one was wrong as most of the score came from WO which requires Z to precede it, and there was no Z in the message. Apart from this fact the comparison was about evens, although if we take into account the fact that there was no better score it would be better. [We have already had a case of this kind of thing in connection with Vigenere; if the various positions are a priori equally likely and the factors are $f_1, f_2 \dots f_N$ then the ~~value~~ probability of the for the rth alternative is better than $\frac{f_r/N}{1+f_r/N}$ $f_r/\Sigma f_i$]

I doubt it SW.

S	F -22	E -34
A	+37 C +37	C -35
T	U 33	T -8
T	I -1	U -40
T	E -2	S -15
W	B -49	H -40
S -10	O -44	L -43
F -31	E -59	E -38
A -11	-22 Y -51	S
S +16	Q	+7 B
T -6 -6	H +8	F
A +10	G	B
U	-26 T -4	-38 R -6
T	J -4	N
E -22	T +4	D
E -24	E +6	X -40
A -29	E -32	G
I -25	F -35	N
E -27	I +1	J -25
U -23	E -52	H -42
F -43	-11 O -14	U
H	R	A
W -53	-13 T -19	N
T 38 -22	A -26	W
J 60	R -45	R
T -52 -25	U -4	
D 37	R	
D -44	N -54	
G -45	L 33	
C +2	N 16	
N -56	N	
I -22	5 N	
T -11 15	-4 N -36	
S -19	A +25	
E -18 -17	I -40	
F	E	
C	R	
U		
I		
E	S	

Fig 7. Scoring the matching of columns in a simple transposition.

Correct matchings ringed.

(Semi-)

Rigorous deduction of the formula (A). (This is something of a digression).

The probability of a piece of plain language coinciding where necessary with the data (D) is

$$p_{\beta_1}\ \tau_{n_2,\beta_1\beta_2}\ \tau_{n_3,\beta_2\beta_3}\ \cdots\ \tau_{n_M,\beta_{M-1}\beta_M}$$

where $\tau_{n,\alpha\beta}$ is

$$\sum_{\eta_1\eta_2\cdots\eta_n} q_{\alpha\eta_1}\, q_{\eta_1\eta_2} \cdots q_{\eta_n\beta}$$

since

$$\sum_{\eta_1\cdots\eta_{n_1}} p_{\eta_1}\, q_{\eta_1\eta_2} \cdots q_{\eta_{n_1}\beta_1} = p_{\beta_1}$$

we can put

$$\tau_{n,\alpha\beta} = (Q^{n+1})_{\alpha\beta}$$

where Q is the matrix whose $\alpha\beta$ coefficient is $q_{\alpha\beta}$: The formula (A) would then be accurate if we could say that for $n>0$, $(Q^{n+1})_{\alpha\beta} = p_\beta$. This is not true, but it is true that except for very special values for $q_{\alpha\beta}$, $(Q^n)_{\alpha\beta} \to p_\beta$ as $n \to \infty$, and the convergence is rather rapid. To prove this I shall assume that the eigenvalues of Q are all different in modulus. In this case we can find a matrix U with unit determinant, such that $U^{-1}QU$ is in diagonal form

$$M = U^{-1}QU = \begin{pmatrix} \mu_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & 0 & \cdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & & 0 & \mu_{26} \end{pmatrix}$$

since $QU = UM$ we have

$$\sum_\gamma q_{\alpha\gamma} u_{\gamma\beta} = \sum_\varepsilon u_{\alpha\varepsilon} m_{\varepsilon\beta}$$

i.e.

$$\sum_\gamma q_{\alpha\gamma} u_{\gamma\beta} = \mu_\beta u_{\alpha\beta}$$

that is, for each β, $u_{\alpha\beta}$ provides a solution of

$$\sum_\gamma q_{\alpha\gamma} l_\gamma = \mu l_\alpha \qquad (E)$$

with $\mu = \mu_\beta$. Conversely if we have any solution of (E) then $\mu = \mu_\theta$, $l_\alpha = k u_{\alpha\theta}$ for some k, θ and all α, for as U is non singular we can find numbers c_γ such that $l_\alpha = \sum_\gamma u_{\alpha\gamma} c_\gamma$ for all α, and then substituting in (E) we get

$$\sum_{\gamma,\delta} q_{\alpha\gamma} u_{\gamma\delta} c_\delta = \mu \sum_\delta u_{\alpha\delta} c_\delta$$

i.e.

$$\sum_\delta (\mu_\delta c_\delta - \mu c_\delta) u_{\alpha\delta} = 0$$

which, since U is non singular implies

$$\mu = \mu_\delta \text{ or } c_\delta = 0 \quad \text{for all } \delta.$$

As the numbers $\mu_1, \ldots, \mu_{26}$ are all different there is only one value θ of δ for which $\mu = \mu_\delta$ and so $l_\alpha = c_\theta u_{\alpha\theta}$ all α.

Now putting $l_\alpha = 1$ for all α we see that one member of the series $\mu_1 \ldots \mu_{2b}$ is 1, for (E) is certainly satisfied. I shall prove that the remaining eigenvalues satisfy $|\mu| \leq 1$. We first prove that if $\mu \neq 1$ then $\sum p_\alpha l_\alpha = 0$. This follows by multiplying (E) on each side by p_α and summing. Since $q_{\alpha\beta} = \frac{p_{\alpha\beta}}{p_\alpha}$ and $\sum_\alpha p_{\alpha\beta} = p_\beta$ we get

$$\sum_{\alpha\gamma} q_{\alpha\gamma} l_\gamma = \sum p_\gamma l_\gamma = \mu \sum p_\alpha l_\alpha$$

which implies $\mu = 1$ or $\sum p_\alpha l_\alpha = 0$. ~~Let l_α satisfy (E) with $\mu \neq 1$. Then $\sum p_\alpha l_\alpha = 0$ and therefore $\Re\, l_\sigma < 0$ for some σ.~~ Next we show that each μ for which $|\mu| > 1$ is real and positive. Let l_α satisfy (E) with $|\mu| > 1$; then the eigenvalue for $\bar{l}_\alpha$ is $\bar{\mu}$ and so

$$\sum_\beta (Q^r)_{\alpha\beta} \left(1 + \varepsilon(l_\beta + \bar{l}_\beta)\right) = 1 + 2\varepsilon \Re\, \mu^r l_\alpha$$

If $\varepsilon > 0$ has been chosen so small that ~~thexRxHxH~~ $\Re\, \varepsilon l_\beta > -\frac{1}{2}$ all β then the L.H.S. is positive for the coefficients in the matrix are positive, whereas the R.H.S. is negative for suitably chosen r, unless $l_\alpha = 0$. If now $\mu > 1$ we may take it that l_α is real for each α. As it must satisfy $\sum p_\alpha l_\alpha = 0$ it is negative for some α, but then

$$\sum_\beta (Q^r)_{\alpha\beta} (1 + \varepsilon l_\beta) = 1 + \varepsilon \mu^r l_\alpha$$

and if ~~the~~ ε is chosen so that ~~thexk~~ $1 + \varepsilon l_\beta > 0$ all β the L.H.S. is positive whereas the R.H.S. is negative for sufficiently large r. All the eigenvalues therefore satisfy $|\mu| \leq 1$.

as the eigenvalues are all different in modulus this means that $|\mu| < 1$ except for one value of μ. Then as $r \to \infty$, M^r tends to a matrix which has only one element different from 0, and that a 1 on the diagonal, say in position $\sigma\sigma$. Calling this matrix X_σ the series of matrices Q^r tends to the limit $U^{-1}X_\sigma U$. This matrix is the one and only one Y which satisfies $YQ = Y$, $Y^2 = Y$, $Y \neq 0$ and is therefore the one whose $\alpha\beta$ coefficient is p_β.

There is another probability problem that arises in connection with simple transposition. With a message of length L, and a key length of K what is the probability that the mth letter will be at the bottom of a column? Let D be the length of the short columns i.e. $D = [L/K]$, and let $E = L - DK$. Then if the mth letter is at the bottom of the wth column we must have $\frac{m}{D+1} \leq w \leq m/D$, and there will be $(D+1)w - m$ short and $m - Dw$ long columns amongst these first w columns. There are $\binom{w}{m-Dw}\binom{K-w}{E-m+Dw}$ ways in which the short and long columns can be arranged consistently with this, and altogether $\binom{K}{E}$ ways in which the columns can be arranged, so that the probability of the mth letter being at the bottom of a column is

$$\sum_{\frac{m}{D+1} \leq w \leq m/D} \binom{w}{m-Dw}\binom{K-w}{E-m+Dw} \Big/ \binom{K}{E}$$

There will normally be very few terms in the sum. Let us take the case of a message of length 133 and consider the 45th letter, assuming the key length is between 10 and 20 (inclusive). i.e. $L = 133$, $m = 45$

$K = 10$, $D = 13$, $E = 3$, $m/D+1 = 3+$ $m/D = 3+$ no terms

$K = 11$, $D = 12$, $E = 1$, $m/D+1 = 3+$ $m/D = 3+$ no terms

$K = 12$, $D = 11$, $E = 1$, $m/D+1 = 3+$ $m/D = 4+$ ~~no terms~~.

~~only term $w = 4$ giving prob^y~~

~~hence~~, only term $w = 4$ giving $m - Dw = 1$ & prob^y $\binom{4}{1}\binom{8}{0}/\binom{12}{1}$

$= 4/12$

$K=13,\ D=10,\ E=3 \quad m/D+1 = 4+,\ m/D = 4+$, no terms

$K=14,\ D=9,\ E=7 \quad m/D+1 = 4+,\ m/D = 5$ only term $w=5$, $m-Dw=0$

$$\text{prob}^y \quad \binom{5}{0}\binom{9}{7}\Big/\binom{14}{7} = 3/286 = .0105$$

$K=15,\ D=8,\ E=13 \quad m/D+1 = 5,\ m/D = 5+$ only term $w=5$, $m-Dw=5$

$$\text{prob}^y \quad \binom{5}{5}\binom{10}{8}\Big/\binom{15}{13} = 3/7 = .428$$

$K=16,\ D=8,\ E=5 \quad m/D+1 = 5+\ \ m/D = 5+$ only term $w=5$ $m-Dw=5$

$$\text{prob}^y \quad \binom{5}{5}\binom{11}{0}\Big/\binom{16}{5} = 1/4368 = .000229$$

$K=17,\ D=7,\ E=14 \quad m/D+1 = 5+\ \ m/D = 6+$ only term $w=6$, $m-Dw=3$

$$\text{prob}^y = \binom{6}{3}\binom{11}{14}\Big/\binom{17}{14} = 1/34 = .0307$$

$K=18,\ D=7\ \ E=7 \quad m/D+1 = 5+\ \ m/D = 6+$ only term $w=6$ $m-Dw=3$

$$\text{prob}^y = \binom{6}{3}\binom{12}{4}\Big/\binom{18}{7} = \frac{4950}{15912} = .311$$

$K=19,\ D=7\ \ E=0 \qquad \text{prob}^y = 0$

$K=20,\ D=6,\ E=13,\ \ m/D+1 = 6+,\ m/D = 7+$ only term $w=7$, $m-Dw=3$

$$\text{prob}^y = \binom{7}{3}\binom{13}{4}\Big/\binom{20}{7} = \frac{35\times 143}{15504} = .323$$